



TITLE:

Flexibility of Tertiary Structures of Proteins :  
Lysozyme and Myoglobin (Commemoration  
Issue Dedicated to Professor Waichiro Tsuji  
On the Occasion of his Retirement)

AUTHOR(S):

Oobatake, Motohisa; Ooi, Tatsuo

---

CITATION:

Oobatake, Motohisa ...[et al]. Flexibility of Tertiary Structures of Proteins : Lysozyme and Myoglobin (Commemoration Issue Dedicated to Professor Waichiro Tsuji On the Occasion of his Retirement). Bulletin of the Institute for Chemical Research, Kyoto University 1974, 52(2): 456-463

ISSUE DATE:

1974-09-13

URL:

<http://hdl.handle.net/2433/76551>

RIGHT:

## Flexibility of Tertiary Structures of Proteins: Lysozyme and Myoglobin

Motohisa OOBATAKE and Tatsuo OOI\*

*Received April 16, 1974*

Flexibility in Lysozyme and Myoglobin was examined by simple methods, one calculating displacements of C $\alpha$  atoms due to small variation of dihedral angle  $\varphi$  (N-C $\alpha$  bond) and  $\psi$  (C $\alpha$ -C' bond), and the other, first derivatives of total energy by using a tentative repulsive potential between two residues. Since a large displacement of a distance in the negative direction (negative maximum displacement) for an interacting pair, or a large value of first derivative represents occurrence of atomic collision, we can infer the flexibility of each residue. The results show that the flexibility for  $\varphi_k$  is almost identical as for  $\psi_{k-1}$  in the opposite direction and terminal regions are more flexible than the rest in the molecule, the range being wider in C-terminal than in N-terminal for both proteins.

### INTRODUCTION

A number of three dimensional structures of proteins have been determined by X-ray crystallography<sup>1)</sup> during the last decade. Since coordinates of atoms thus determined are not accurate enough to compute an internal energy of a protein, several attempts to obtain an energetically stable conformation, *e.g.*, energy minimization techniques have been applied to proteins.<sup>2-6)</sup> Such refinement of coordinates, however, requires much computing time for the calculation of total energy of the protein by summing up all pairwise interactions. The use of coordinates of heavy atoms (C, O, and N), on the other hand, make it possible to examine such properties as flexibility of the stable structure of a protein around the equilibrium, which seems to be related closely to folding and unfolding process, stability of the protein, catalytic reaction mechanism, and allosteric effect, *etc.*

The conformation of a protein may be described as a function of dihedral angles  $\varphi$  (about N-C $\alpha$  bond) and  $\psi$  (about C $\alpha$ -C' bond) along the main chain, under the assumptions of fixed bond lengths and bond angles, and the standard geometry of a peptide. For the calculation of flexibility about a residue it is necessary to compute energy changes due to variation of  $\varphi$  and  $\psi$  from the equilibrium values. Here simple methods for saving the computing time are introduced, considering the nature of interaction energy. First, a change in the distance of a residue pair which is in the interacting range (*e.g.*, within 10 Å) will give informations about the change in energy whether the interaction becomes repulsive or not, because atoms at equilibrium are in an adequate distance near energy minimum favorable for attraction. Computations of such displacements are easy to be done. Second, we shall use a tentative repulsive potential between amino acid residues, instead of calculating all the atomic interactions. These methods will simplify the procedure to estimate flexibility in a protein.

\* 大島玄久, 大井竜夫: Laboratory of Physical Chemistry of Enzyme, Institute for Chemical Research, Kyoto University, Uji, Kyoto.

## METHODS

## 1. Negative Maximum Displacements

Intramolecular energies of a protein molecule are classified into non-bonded, electrostatic, hydrogen bond, torsional, and hydrophobic energy, among which most important one at present is the repulsive part of the non-bonded energy, since atomic collisions do not occur in any conformation. The repulsive part of the non-bonded energy is inversely proportional to 12th power of a pairwise distance, so that the displacement towards a smaller distance of an atomic pair results in a large energy increase when the pair is near potential minimum. We simplify the system by paying attention to the repulsive energy such that movements of every atoms on a residue are replaced to a movement of  $C^\alpha$  atom of the residue. That is, changes in energy between residues are estimated by a degree of a displacement of atomic pairs which yields repulsive energy.

According to the above consideration, it is not necessary to compute all the changes in atomic pair distances, since an atomic pair of large separation has no contribution to the stability of the conformation. We have chosen the limit of the distance as 10 Å between  $C^\alpha$  atoms, beyond which almost no effect is expected to exist. We adopted conformations,  $A_f$  for Lysozyme and  $K_f$  for Myoglobin in the previous paper<sup>3)</sup> as stable ones which are expressed in terms of dihedral angles ( $\varphi_1\psi_1\dots\varphi_N\psi_N$ ) where  $N$  is the residue number. The distance between  $C^\alpha$  atoms of  $i$ -th and  $j$ -th residues,  $R_{ij}^s$ , is calculated easily from the set of  $\varphi$  and  $\psi$  by using transformation matrices. When a dihedral angle of  $k$ -th residue, *e.g.*,  $\varphi_k$ , is rotated by a small amount, namely  $1^\circ$ , mutual distance  $R_{ij}^{k_{ij}}$  ( $i,j=1\sim N$ ) changes from  $R_{ij}^s$  by  $\Delta R_{ij}^{k_{ij}}$ , that is,

$$\Delta R_{ij}^{k_{ij}} = R_{ij}^{k_{ij}} - R_{ij}^s \quad (1)$$

Here (ij) pairs were selected on the basis that  $R_{ij}^s$  is smaller than 10 Å.  $(\Delta R_{ij}^{k_{ij}})_{\max}$  is the maximum in absolute value among negative displacements of  $\Delta R_{ij}^{k_{ij}}$ . The maximum value of the negative displacement  $(\Delta R_{ij}^{k_{ij}})_{\max}$  indicates a possibility of occurrence in severe atomic collision for the (ij) pair, at a small variation of  $\varphi_k$  or  $\psi_k$ . Since the increment can be taken in positive or negative direction,  $D_k$  is assigned as the lowest one in absolute value out of  $(\Delta R_{ij}^{k_{ij}})_{\max}$  in both directions, because a lower value of  $(\Delta R_{ij}^{k_{ij}})_{\max}$  would correspond to less repulsive collision when compared with the changes of the conformation produced in both directions, *e.g.*, the direction having a lower value of  $(\Delta R_{ij}^{k_{ij}})_{\max}$  would be favorable than the other direction.

The flexibility occurring in a medium range interaction is estimated by the same procedure, taking pair interactions in the range into consideration, *e.g.*, the interactions involved within ten residues before and behind  $k$ -th residue. That is, search of  $\Delta R_{ij}^{k_{ij}}$  in Eq. (1) is made over  $i=1\sim N$  and  $j=i-10\sim i+10$  except for both terminal regions.

## 2. A Repulsive Energy Function

The repulsive energies between residues are plotted against the separation of the residues from the data in the computations of energies of Lysozyme and Myoglobin.<sup>3)</sup> From these plots shown in Fig. 1, an empirical potential as a function of the distance between  $C^\alpha$  atoms (as a representative distance between residues) may be assigned in the form of  $A/r_{ij}^{10}$  where  $r_{ij}$  is the distance between  $C^\alpha$  atoms of  $i$ -th and  $j$ -th residue. The constant  $A$  was chosen as  $2.8 \times 10^7$  Kcal Å<sup>10</sup>/mol to fit the plots in Fig. 1. Then a derivative of

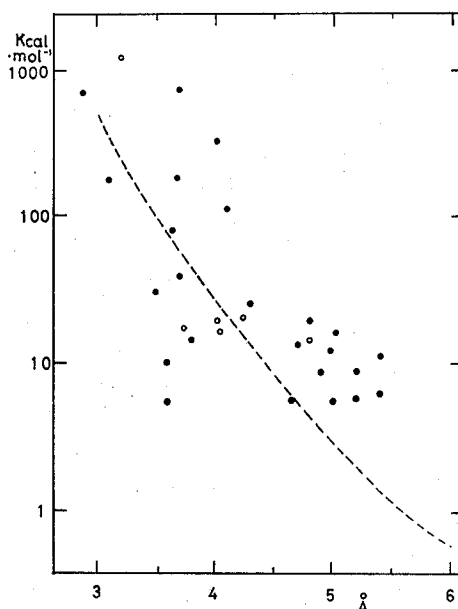


Fig. 1. Plots of nonbonded energy *vs* pair distance between C $\alpha$  atoms. The dashed line represents a repulsion energy function. The open and closed circles represent the data for Myoglobin and Lysozyme in the previous paper, respectively.

the total energy by the use of the above potential with respect to  $k$ -th dihedral angle,  $\varphi_k$ , may be written as

$$G_k = -\frac{\partial}{\partial \varphi_k} \left( \sum_{i=1}^N \sum_{j=1}^i \frac{A}{r_{ij}^{10}} \right) \\ = -\sum_{i=1}^N \sum_{j=1}^i \left( \frac{10A}{r_{ij}^{11}} \frac{\partial r_{ij}}{\partial \varphi_k} \right) \quad (2)$$

The derivative  $\partial r_{ij}/\partial \varphi_k$  can be obtained easily as follows; the deviation of vector  $r_{ij}$  is represented by a vector product of  $\mathbf{c}_k$ , a vector whose components are direction cosines of the bond N $_k$ -C $\alpha_k$  (for  $\varphi_k$ ) and  $\mathbf{b}_{kij}$ , a vector connecting C $\alpha$  atom of  $k$ -th residue to C $\alpha$  atom of  $j$ -th residue ( $i > j$ ), or

$$d\mathbf{r}_{ij} = (\mathbf{c}_k \times \mathbf{b}_{kij}) d\varphi_k \quad (3)$$

Combining Eqs. (2) and (3), we have

$$G_k = -\sum_{i=1}^N \sum_{j=1}^i \frac{10A}{r_{ij}^{12}} [\mathbf{r}_{ij} \cdot (\mathbf{c}_k \times \mathbf{b}_{kij})] \quad (4)$$

A change of  $\varphi_k$  from positive direction to negative results only in the change of sign of  $G_k$  in Eq. (4).

All the computations were carried out with FACOM 230-60 at the computing center of Kyoto University.

## RESULTS

## 1. Estimate of Flexibility by the Use of Displacement Technique

Computed values of  $D_k$  obtained by a small rotation of  $\varphi_k$  and  $\psi_k$  plotted against the residue number for Lysozyme and Myoglobin are shown in Figs. 2 and 3, respectively. Distribution of  $D_k$  vs. the residue number resembles to plots of distances from the center of gravity to every atoms except for N and C terminal regions. It is notable that  $D_k$  for  $\varphi_k$  is approximately the same as  $D_{k-1}$  for  $\psi_{k-1}$ . This is due to the planar structure of a peptide where the bonds  $C_{k-1}^{\alpha}-C'_{k-1}$  and  $N_k-C_k^{\alpha}$  are almost in parallel with each other. General feature of both patterns of Lysozyme and Myoglobin is that large and small values are repeated alternatively. N and C terminal regions, where  $D_k$  values are relatively small, are different in the number of residues, larger for C terminal region than for N-terminal. Since a large value of  $D_k$  means possibility of atomic collision, terminal regions

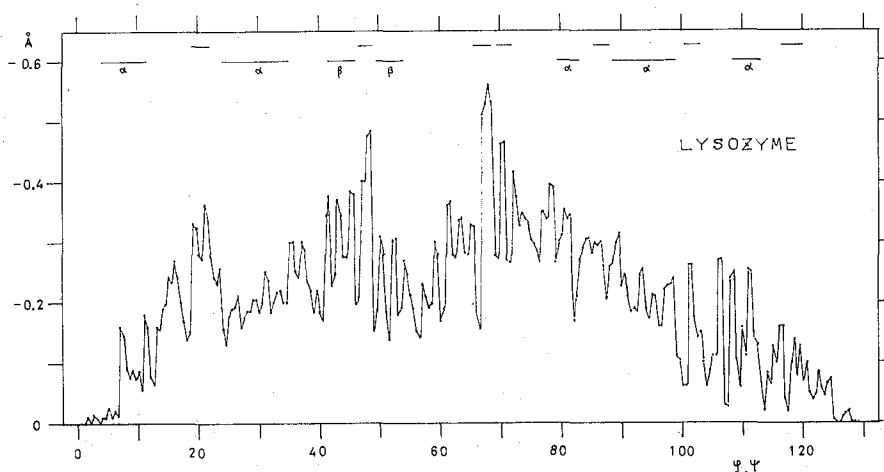


Fig. 2. The negative displacements  $D_k$  due to small deviation of dihedral angles of residues ( $\varphi_1\psi_1$  .....  $\varphi_N\psi_N$ ) are aligned in the sequence of Lysozyme. The upper bars shown in the top of the figure represent turn regions, and symbols  $\alpha$  and  $\beta$ ,  $\alpha$ -helix and  $\beta$ -structure regions, respectively.

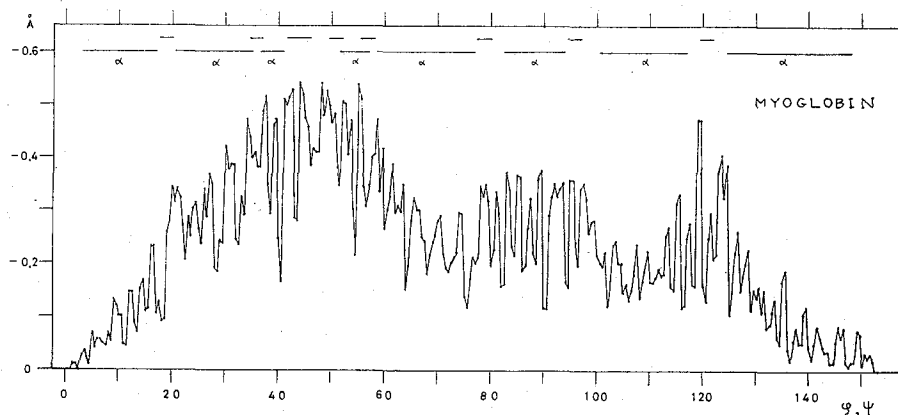


Fig. 3.  $D_k$  plots for Myoglobin as in Fig. 2.

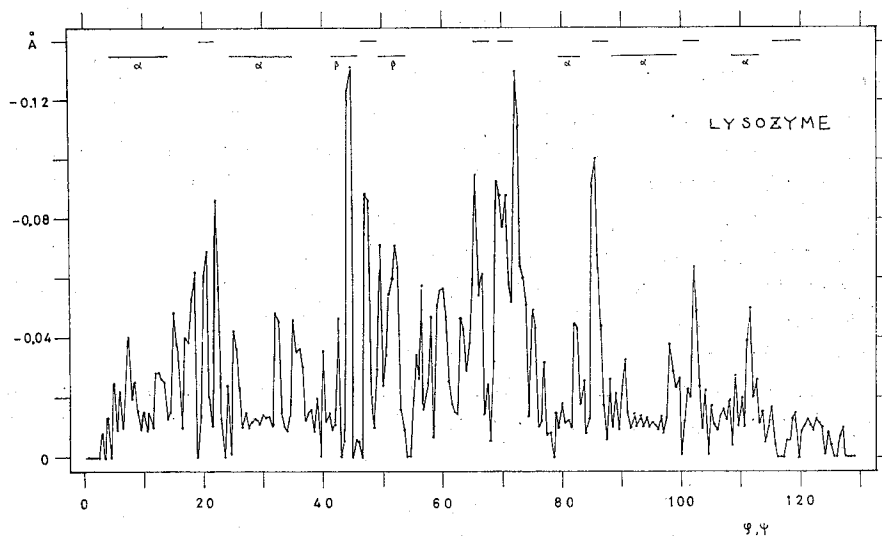


Fig. 4.  $D_K$  for the short range interaction ten residues before and behind a residue is plotted in the sequence ( $\phi_1\psi_1$  .....  $\phi_N\psi_N$ ) for Lysozyme.

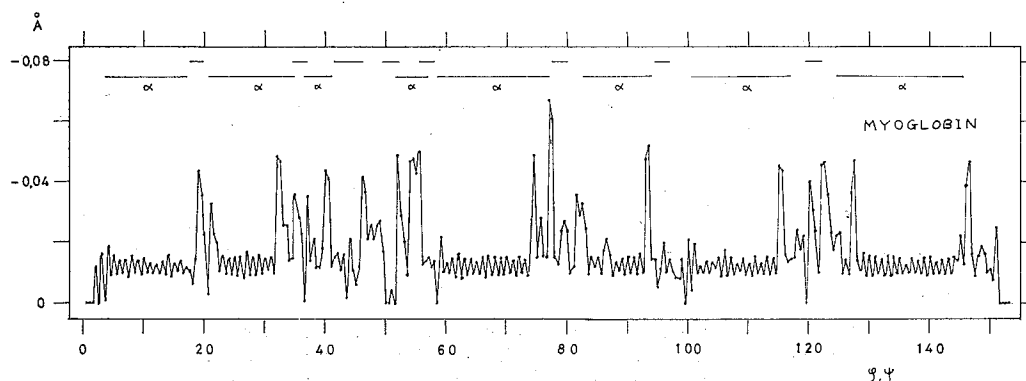


Fig. 5.  $D_K$  plots as in Fig. 4 for Myoglobin.

especially near C-terminals are considered to be more flexible than the other regions. The large values of  $D_K$  appeared generally in turn regions, *e.g.*, residue 44 and 68 for Lysozyme and 44 and 120 for Myoglobin have high  $D_K$  values.

Figures 4 and 5 show the pattern of  $D_K$  for middle range interaction, taking ten residues before and behind a residue. The regular zig-zag pattern corresponds to  $\alpha$ -helix, which is demonstrated clearly in Myoglobin.  $D_K$  values of turn region and both ends of  $\alpha$ -helix are irregularly high and low.

## 2. Estimate of Flexibility by the Use of the Repulsive Energy Function

The results described above are based only on extents of negative displacements of mutual pairs existing within 10 Å, so that the flexibility estimated by this method is quite rough. On the other hand, the similar results shown in Figs. 6 and 7 are more realistic for the estimate of flexibility of each residue, since total energy was taken into account, though a repulsive part only, for the computations. As shown previously,  $D_K$  value for

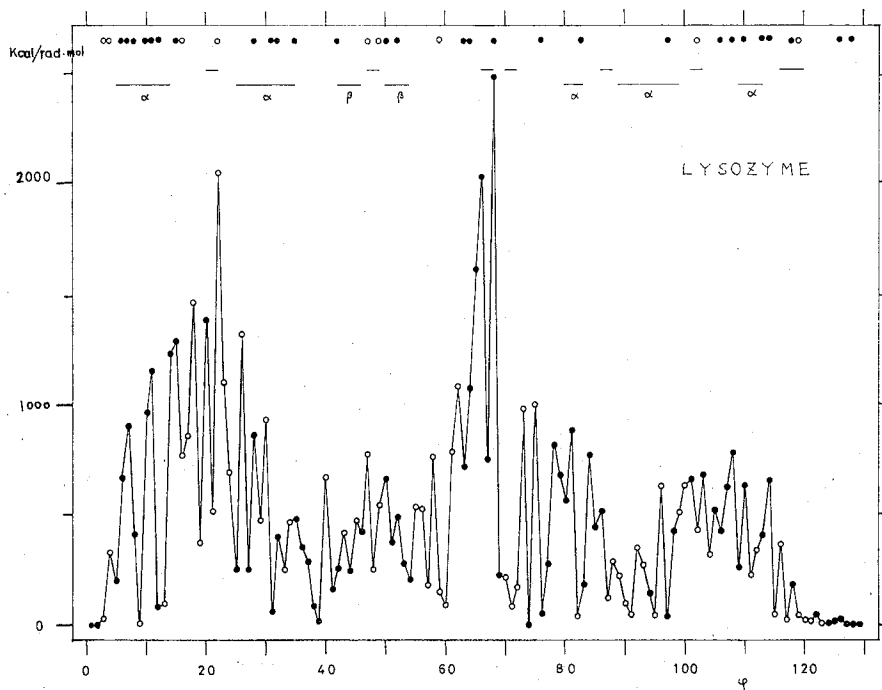


Fig. 6. First derivatives,  $G_k$ , with respect to  $\varphi_k$  are shown in the sequence of the residue number for Lysozyme. Open and closed circles denote negative and positive values of  $G_k$ . The upper open and closed circles represent that all the derivatives of interacting pairs are negative and positive values, respectively.

$\varphi_k$  is almost the same as for  $\psi_{k-1}$ , and, therefore, it is unnecessary to compute a derivative of the total energy using the present potential for both  $\varphi_k$  and  $\psi_{k-1}$ , but only computation about  $\varphi_k$  may be sufficient.

General pattern of  $G_k$  is repetition of high and low values with a period over one to a few residues. The regions which include derivatives higher than 1000 Kcal/mol/rad are those from the residue 10 to 27, and from 61 to 68 for Lysozyme, and 60 and 92 to 96 for Myoglobin. As deduced from the previous results, both terminal regions have low values of  $G_k$ , wider for C-terminal than for N-terminal, and greater in number for Myoglobin than for Lysozyme, suggesting these regions are more flexible than the other parts of the proteins.

Since fluctuation occurs under thermal motion in the system, the approximate amount of  $\Delta\varphi_k$  for  $k$ -th residue would be proportional to  $RT/G_k$ . Simple calculation leads to  $\Delta\varphi \sim 0.04^\circ$  for  $G_k \sim 800$  kcal/mol/rad and  $\Delta\varphi \sim 0.2^\circ$  for  $G_k \sim 150$  kcal/mol/rad. In Table I, the residues which have larger  $G_k$  than 800 Kcal/mol/rad and those smaller than 150 kcal/mol/rad were shown in the primary structure of Lysozyme and Myoglobin, respectively. It is clearly shown that fluctuation near C-terminal regions would be large. The more exact amount of fluctuation,  $\langle \Delta\varphi_k^2 \rangle^{1/2}$ , may be obtained from second derivatives.

Another aspect of the results obtained here is information about directions of  $\Delta\varphi_k$ . Since a derivative  $G_k = \partial E / \partial \varphi_k$  consists of components,  $\partial E_{ij} / \partial \varphi_k$ , we can examine those components for every residues, which are all negative, or negative and positive, or all positive. These signs of direction are demonstrated in top columns of Figs. 6 and 7,

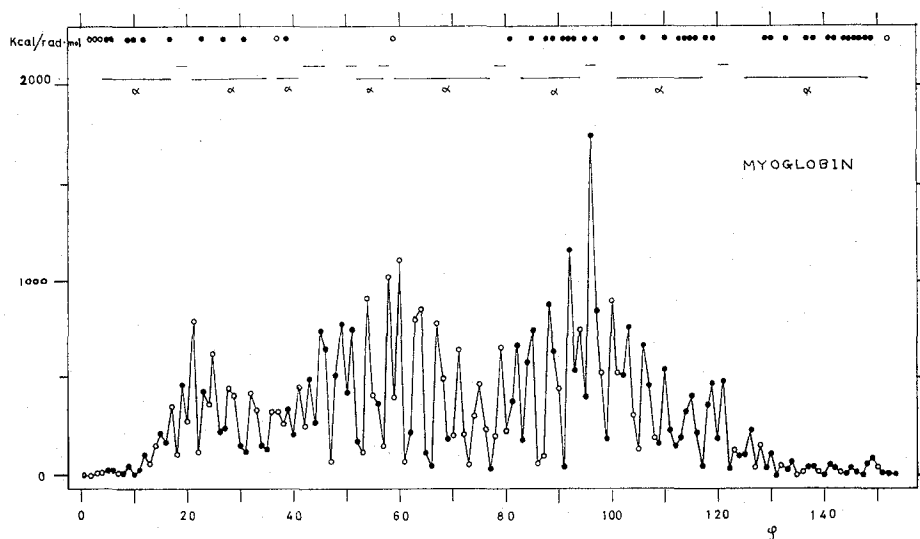
Fig. 7.  $G_k$  plots as in Fig. 6 for Myoglobin.

Table I. Listings of Flexibility in the Sequences.

| <i>Lysozyme</i>  |          |          |          |          |          |          |          |          |          |          |
|------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
|                  | 5        | 10       | 15       | 20       | 25       | 30       |          |          |          |          |
| 1                | <u>K</u> | <u>V</u> | <u>F</u> | G        | R        | C        | <u>E</u> | L        | <u>A</u> | <u>A</u> |
| 31               | <u>A</u> | <u>A</u> | <u>K</u> | <u>F</u> | <u>E</u> | <u>S</u> | <u>N</u> | <u>F</u> | <u>N</u> | T        |
| 61               | R        | W        | W        | C        | N        | D        | G        | R        | T        | P        |
| 91               | <u>S</u> | <u>V</u> | <u>N</u> | <u>C</u> | <u>A</u> | <u>K</u> | <u>K</u> | I        | V        | S        |
| 121              | <u>Q</u> | <u>A</u> | <u>W</u> | <u>I</u> | <u>R</u> | <u>G</u> | <u>C</u> | <u>R</u> | <u>L</u> |          |
| <i>Myoglobin</i> |          |          |          |          |          |          |          |          |          |          |
|                  | 5        | 10       | 15       | 20       | 25       | 30       |          |          |          |          |
| 1                | <u>V</u> | <u>L</u> | <u>S</u> | <u>E</u> | <u>G</u> | <u>E</u> | <u>W</u> | <u>Q</u> | <u>L</u> | V        |
| 31               | <u>R</u> | <u>L</u> | <u>F</u> | <u>K</u> | <u>S</u> | H        | P        | E        | T        | L        |
| 61               | <u>L</u> | <u>K</u> | <u>K</u> | <u>H</u> | <u>G</u> | <u>V</u> | T        | V        | L        | T        |
| 91               | <u>Q</u> | <u>S</u> | <u>H</u> | <u>A</u> | <u>T</u> | <u>K</u> | <u>H</u> | <u>K</u> | <u>I</u> | P        |
| 121              | <u>G</u> | <u>N</u> | <u>F</u> | <u>G</u> | <u>A</u> | <u>D</u> | <u>A</u> | <u>Q</u> | <u>G</u> | <u>A</u> |
| 151              | <u>Y</u> | <u>Q</u> | <u>G</u> |          |          |          |          |          |          |          |

Residues which have  $G_k$  values less than 150 kcal/mol/rad are underlined, and those greater than 800 kcal/mol/rad are shown by dots.

filled circle for positive direction, open circle for negative and not indicated for both direction. That is, movements of the residues are limited in a certain direction depending on the sign. Interestingly only a few residues have all negative signs in Myoglobin.



## DISCUSSION

The methods described here are convenient for rough estimation of flexibility in a protein without computing total energy precisely. The results, however, might be overestimated, since the methods are taken only displacements of C $\alpha$  atoms or derivatives of repulsive potentials into consideration. Nevertheless, the results especially by the use of the repulsive potential would indicate the characteristic feature of flexibility, *e.g.*, more freedom around C-terminal regions than the rest in the molecule.

To evaluate fluctuation in a protein of native state, we have to know the exact location of every atoms and then to compute second derivatives with respect to variables, yielding the degree of correlation  $\langle \Delta\varphi_i \Delta\varphi_j \rangle$  including the amount of fluctuation  $\langle \Delta\varphi_i^2 \rangle$ . Performance of such computations requires much computing time and, therefore, the present procedures would be worthwhile to provide rough informations about the flexibility around each residue.

Suppose unfolding of a protein occurs according to a change in environments, then the passway to unfolding may be along gentlest ascend of the energy surface, or movements of dihedral angles would be inversely proportional to  $G_k$ . In other words, first step of unfolding would be relatively large rotations about residues which have low  $G_k$  values shown in Figs. 6 and 7. Furthermore, it is desirable that components of  $G_k$  have the same sign so that the sense of rotation would be favorable for all the components. The residues which satisfy such conditions are quite few in the middle of the sequences, and locate mainly in the terminal regions. Presumably those residues are candidates of initial movements at unfolding.

This research was supported by the grant from Institute of Fundamental Physics of Kyoto University.

## REFERENCES

- (1) "Cold Spr. Harb. Symp. Quant. Biol". University of Tokyo Press, Vol. 36, 1971.
- (2) K. Nishikawa, T. Ooi, Y. Isogai, and N. Saito, *J. Phys. Soc. Japan*, **32**, 1331 (1972).
- (3) K. Nishikawa and T. Ooi, *ibid.*, **32**, 1338 (1972).
- (4) M. Levitt and S. Lifson, *J. Mol. Biol.*, **46**, 269 (1969).
- (5) P. K. Warne, N. Go, and H. A. Scheraga, *J. Computational Phys.*, **9**, 303 (1972).
- (6) P. K. Warne and H. A. Scheraga, *ibid.*, **12**, 49 (1973).
- (7) K. Nishikawa, M. Oobatake, and T. Ooi, *Bull. Inst. Chem. Res., Kyoto Univ.*, **48**, 102 (1970).